



METIS DATA SCIENCE BOOTCAMP

INTRO TO MACHINE LEARNING



WHAT IS MACHINE LEARNING?

WHAT IS MACHINE LEARNING

“MACHINE LEARNING, A BRANCH OF ARTIFICIAL INTELLIGENCE, IS ABOUT THE CONSTRUCTION AND STUDY OF SYSTEMS THAT CAN LEARN FROM DATA.”

source: http://en.wikipedia.org/wiki/Machine_learning

“MACHINE LEARNING, A BRANCH OF ARTIFICIAL INTELLIGENCE, IS ABOUT THE CONSTRUCTION AND STUDY OF SYSTEMS THAT CAN LEARN FROM DATA.”

source: http://en.wikipedia.org/wiki/Machine_learning

“The core of machine learning deals with representation and generalization...”

- ▶ Representation - extracting structure from data
- ▶ Generalization - making predictions from data

WHAT IS MACHINE LEARNING

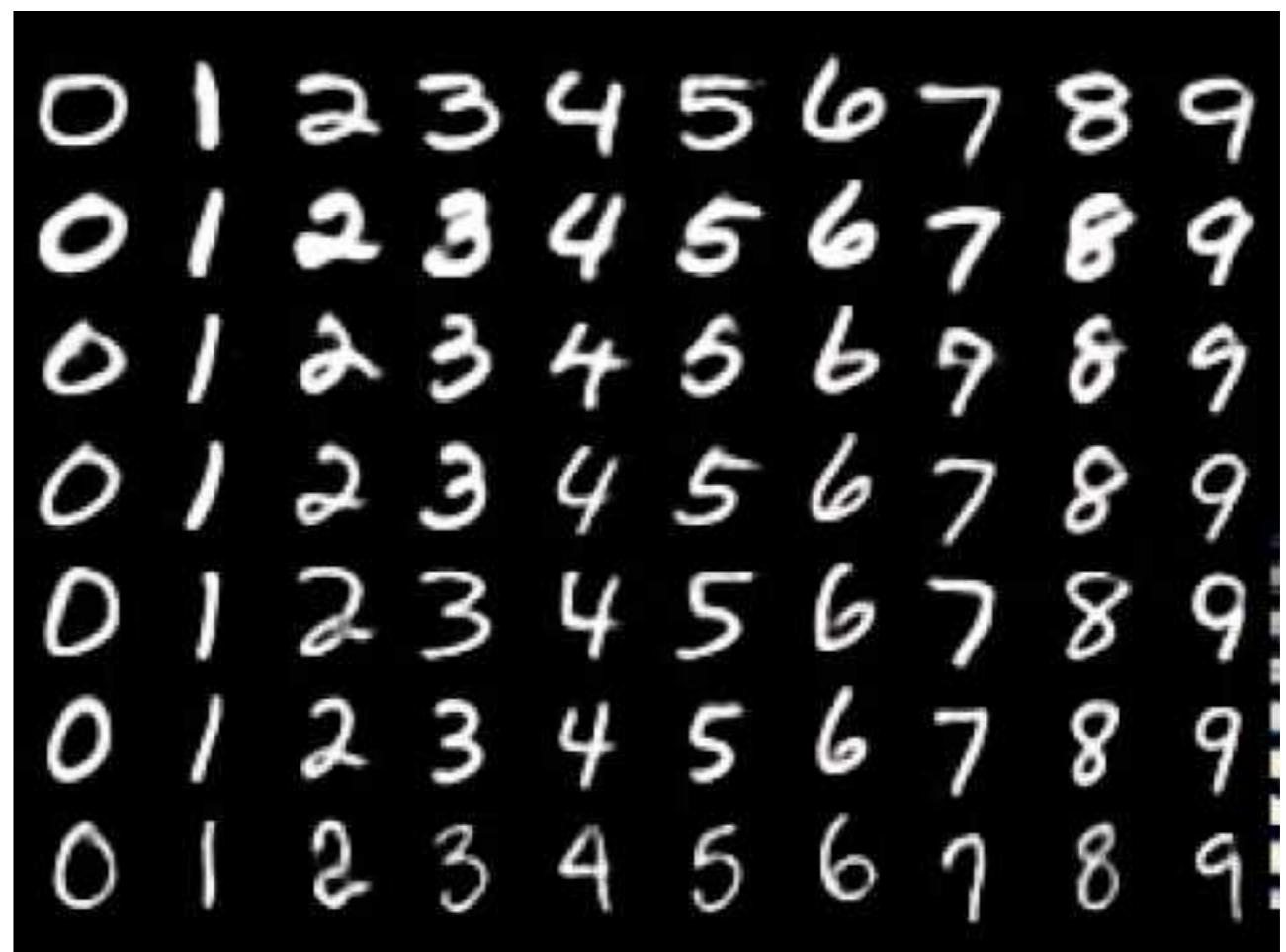
Learning is not about memorization and recollection. It is about **generalizing** conclusions to previously unseen examples.

SUPERVISED LEARNING

- ▶ We have data points where we know the ‘correct’ output. For instance, we know both y and x . We want to learn about the mapping.
- ▶ This is known as learning from ‘labeled’ data.

SUPERVISED LEARNING

Optical Character
Recognition



TYPES OF LEARNING

SUPERVISED LEARNING

**CLICK HERE
TO APPLY TODAY!**



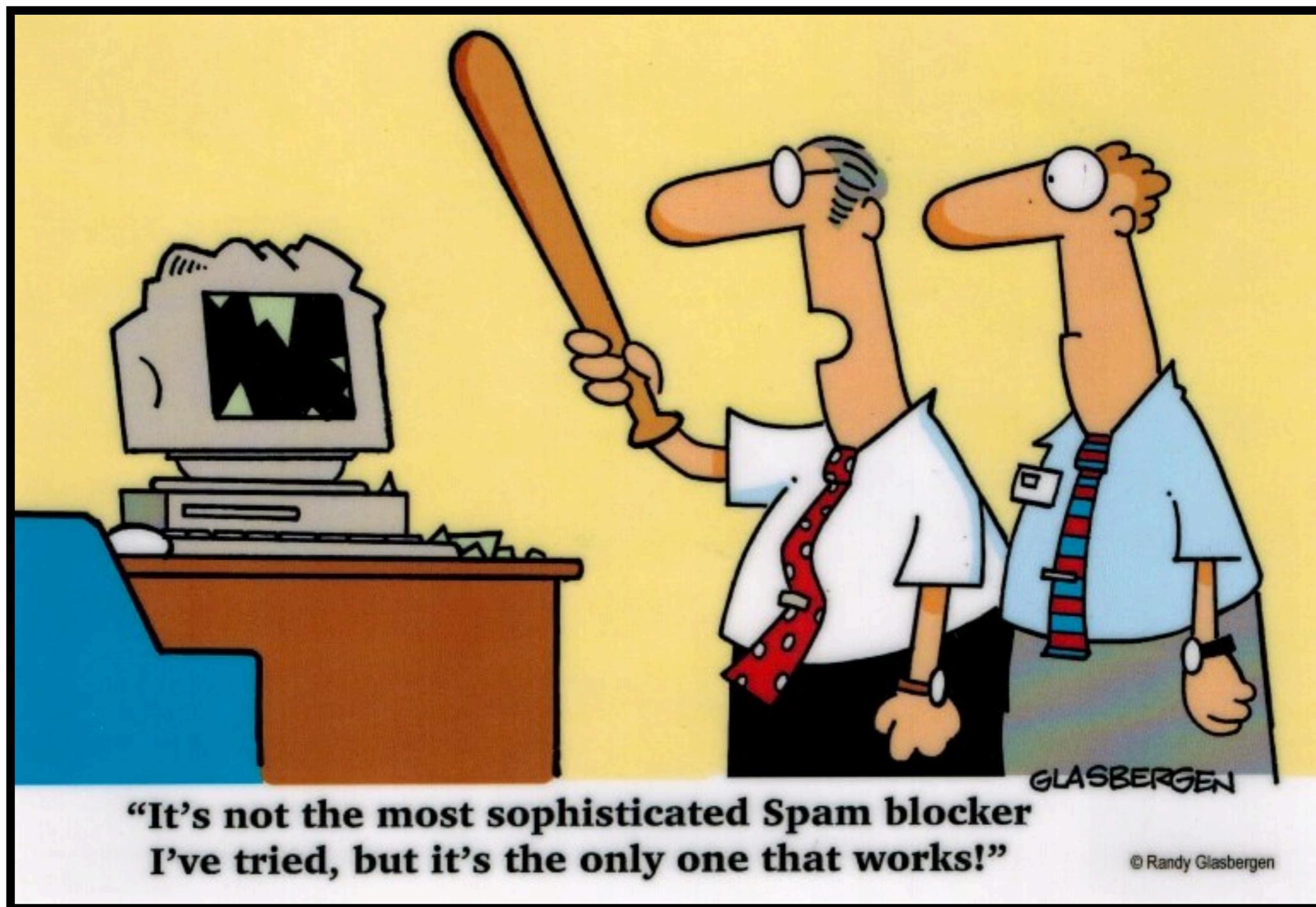
	<i>Client 1</i>	<i>Client 2</i>	<i>Client 3</i>
<i>Age</i>	23	30	19
<i>Gender</i>	<i>M</i>	<i>F</i>	<i>M</i>
<i>Annual salary</i>	\$30,000	\$45,000	\$15,000
<i>Years in residence</i>	3 years	1 year	3 month
<i>Years in job</i>	1 year	1 year	1 month
<i>Current debt</i>	\$5,000	\$1,000	\$10,000
<i>Paid off credit</i>	Yes	Yes	No

TYPES OF LEARNING

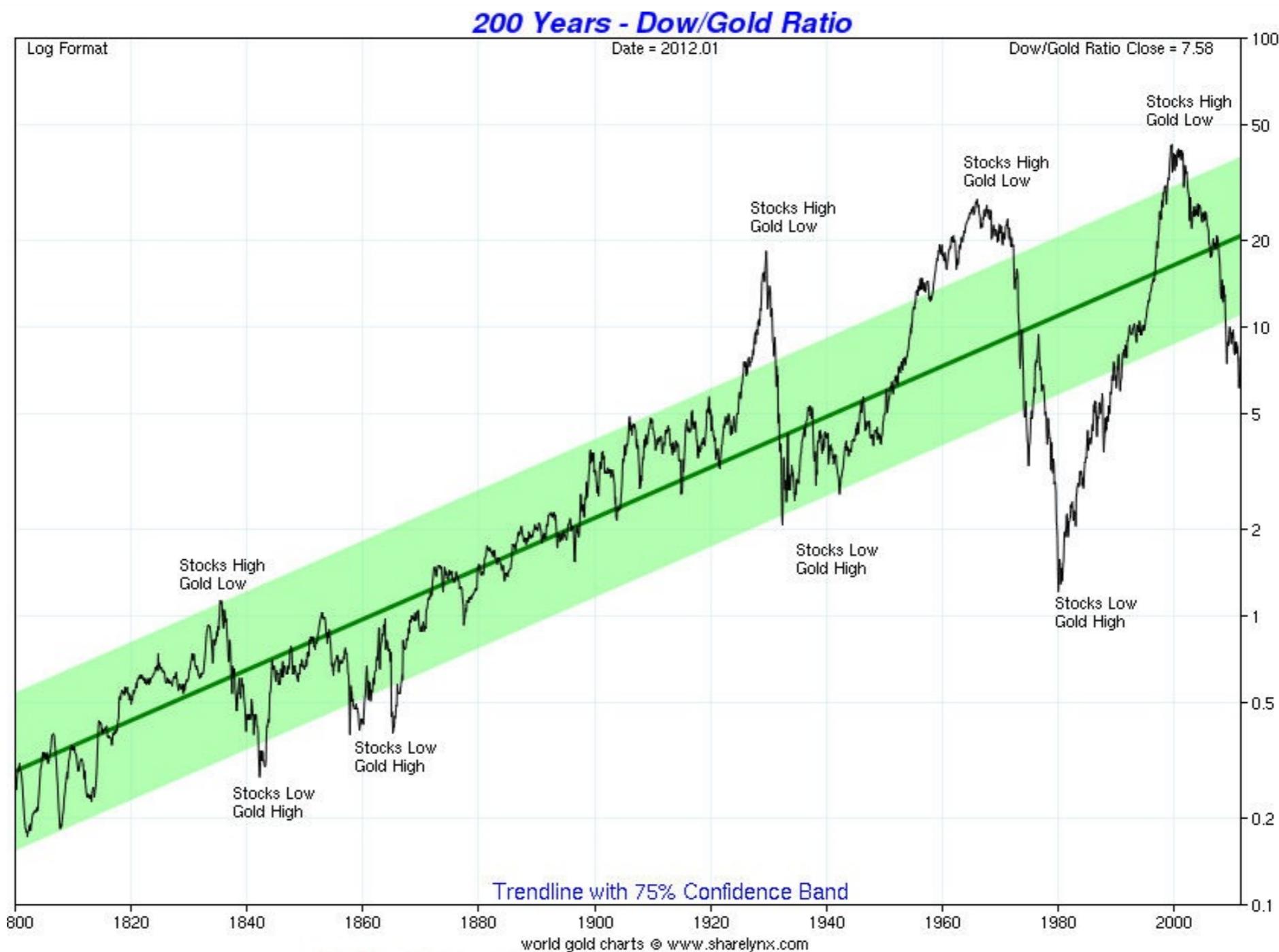
SUPERVISED LEARNING

	<i>Client 1</i>	<i>Client 2</i>	<i>Client 3</i>		<i>Applicant</i>
<i>Age</i>	23	30	19	<i>Age</i>	25
<i>Gender</i>	M	F	M	<i>Gender</i>	M
<i>Annual salary</i>	\$30,000	\$45,000	\$15,000	<i>Annual salary</i>	\$25,000
<i>Years in residence</i>	3 years	1 year	3 month	<i>Years in residence</i>	1 year
<i>Years in job</i>	1 year	1 year	1 month	<i>Years in job</i>	2 years
<i>Current debt</i>	\$5,000	\$1,000	\$10,000	<i>Current debt</i>	\$15,000
<i>Paid off credit</i>	Yes	Yes	No	<i>Credit decision/score</i>	???

SUPERVISED LEARNING



SUPERVISED LEARNING

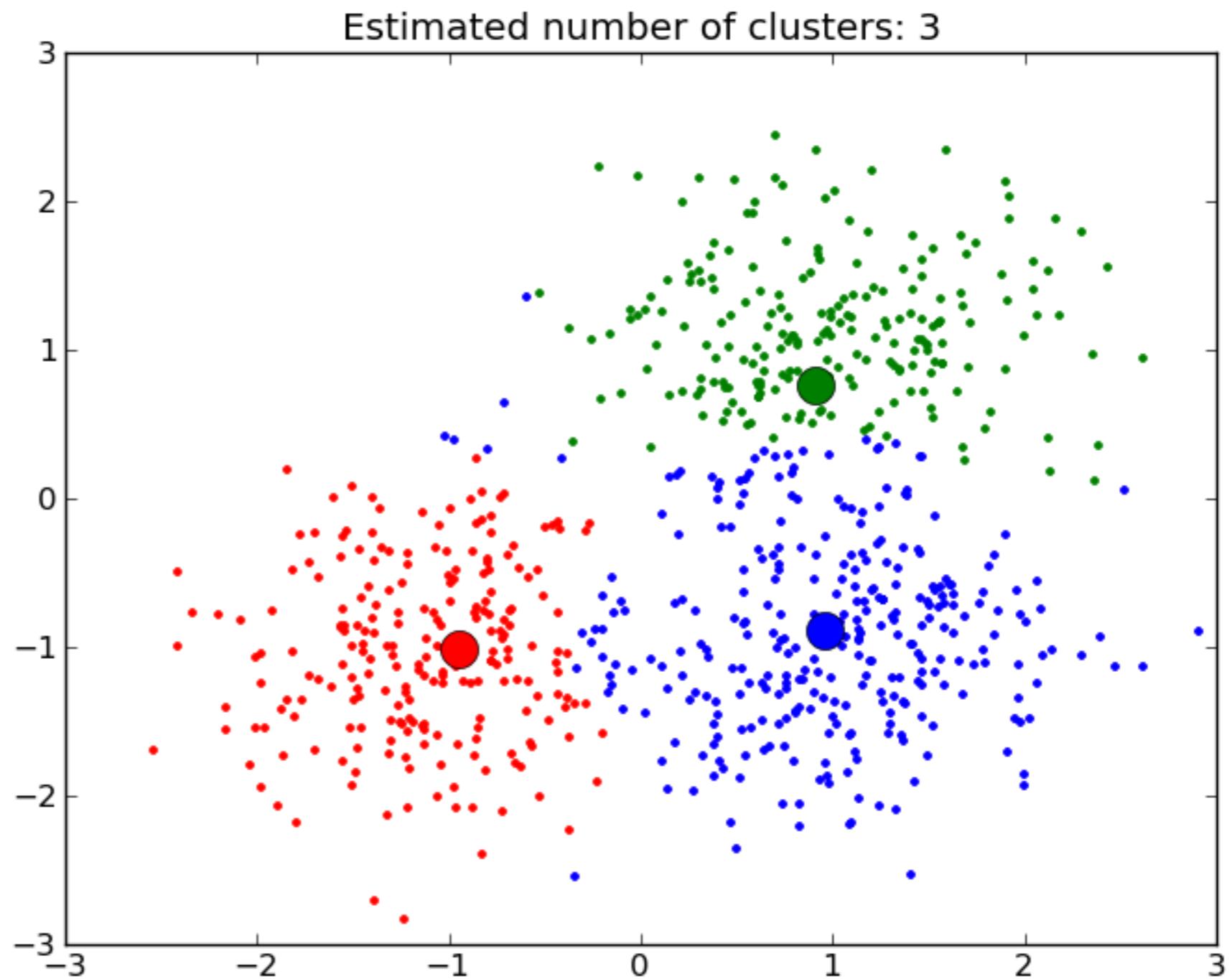


UNSUPERVISED LEARNING

- ▶ Goal is to learn about the underlying structure of the data.
- ▶ No labels to work with.

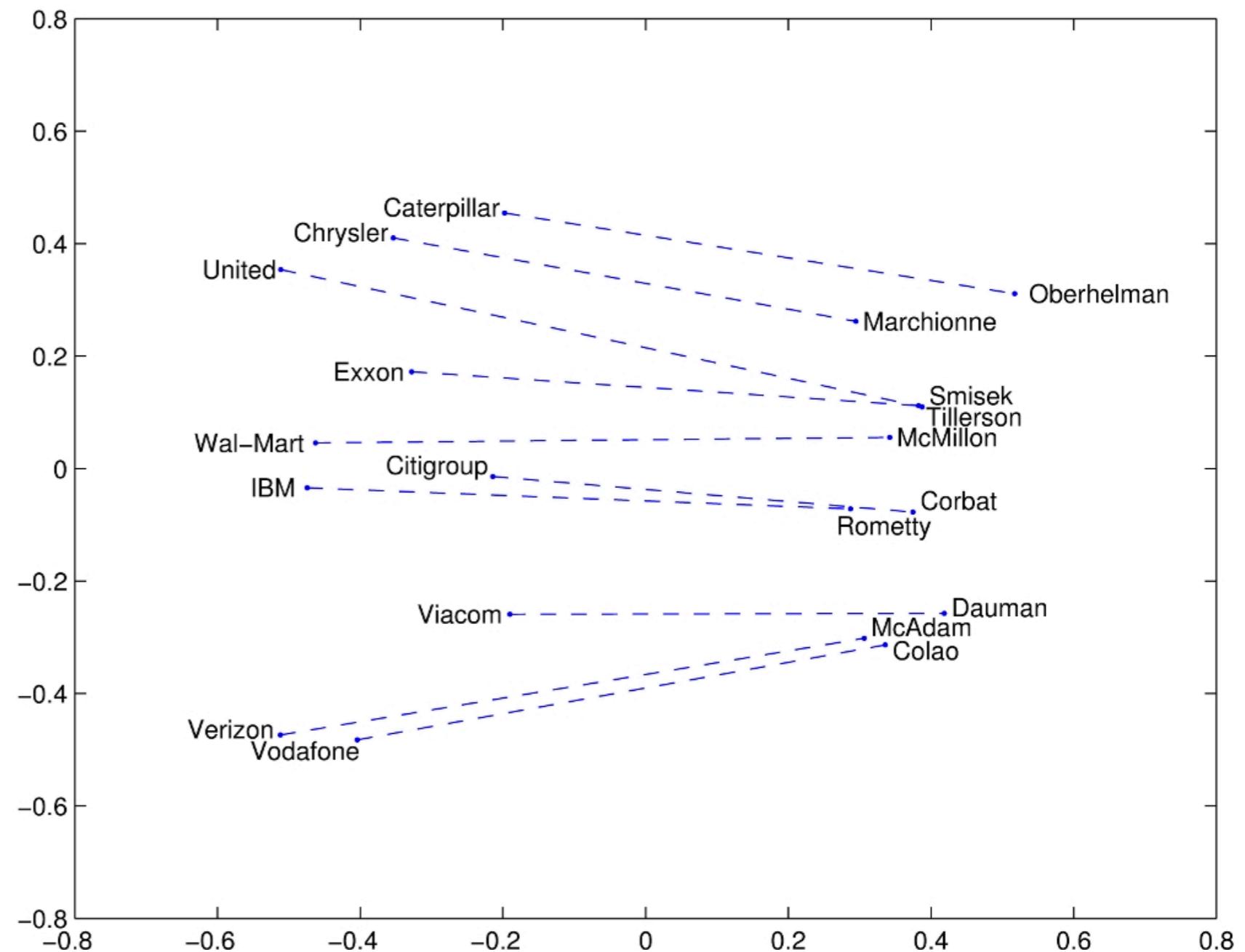
UNSUPERVISED LEARNING

Clustering



UNSUPERVISED LEARNING

Natural
Language
Processing



TYPES OF LEARNING

Supervised

Generalizing/Predicting

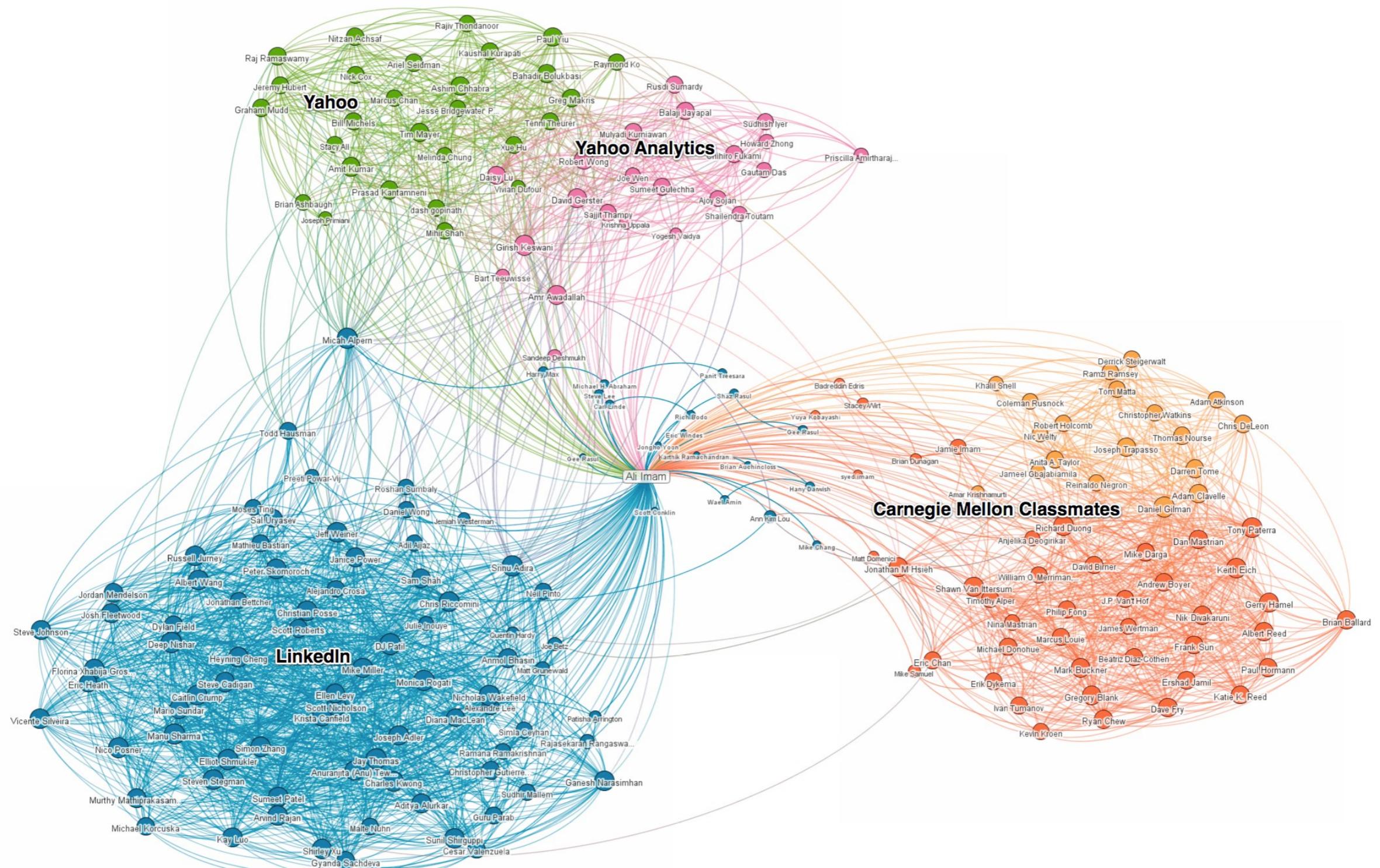
Unsupervised

Extracting Structure



**SUPERVISED
OR
UNSUPERVISED?**

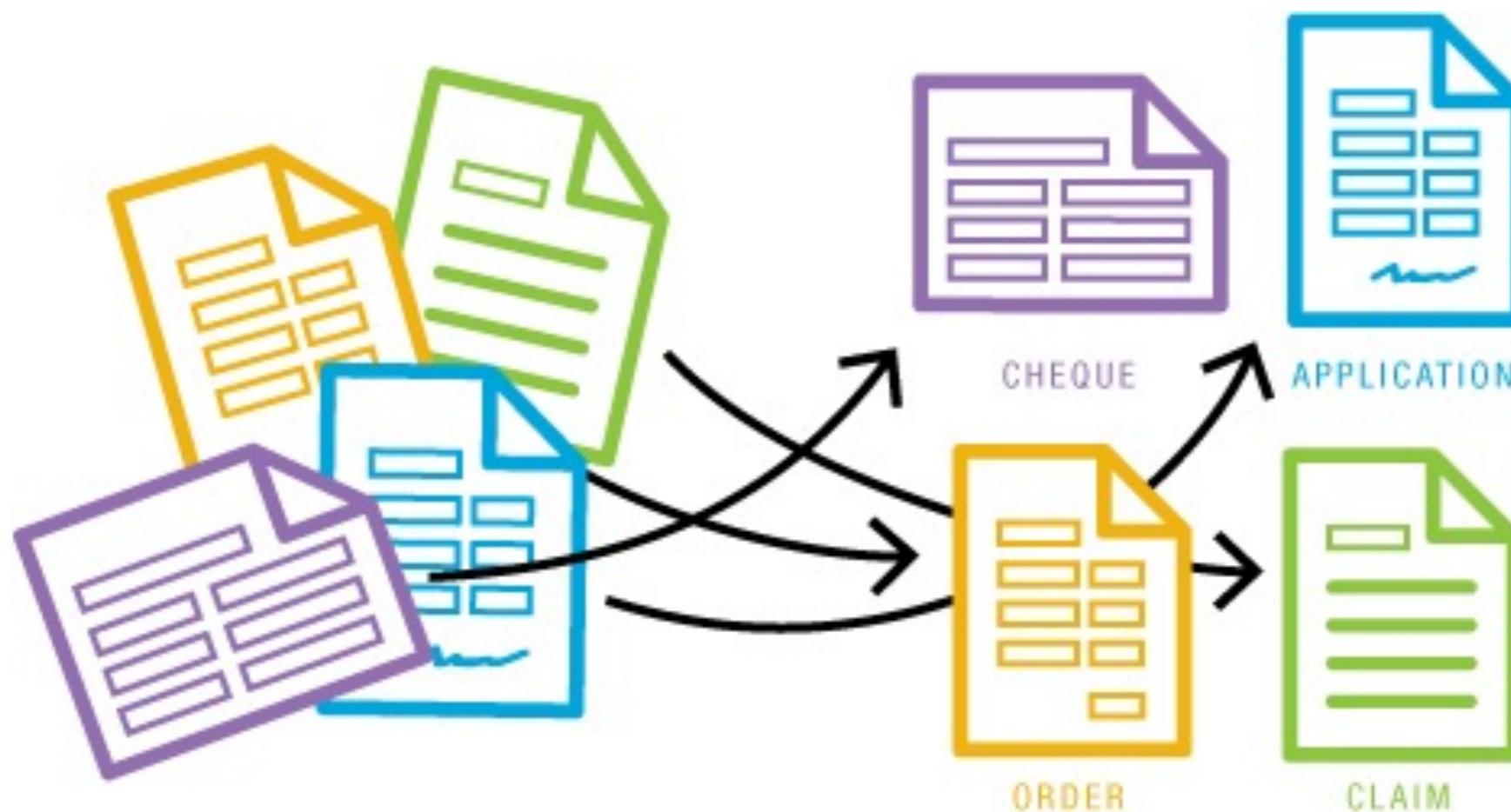
SUPERVISED OR UNSUPERVISED? - COMMUNITY DETECTION



SUPERVISED OR UNSUPERVISED? - HOUSE PRICE PREDICTION



SUPERVISED OR UNSUPERVISED? - DOCUMENT CLASSIFICATION



TYPES OF DATA

Continuous

quantitative

Categorical

qualitative

TYPES OF DATA

Continuous

Height

Speed

Time

Temperature

Price

Categorical

Team

Spam?

Eye Color

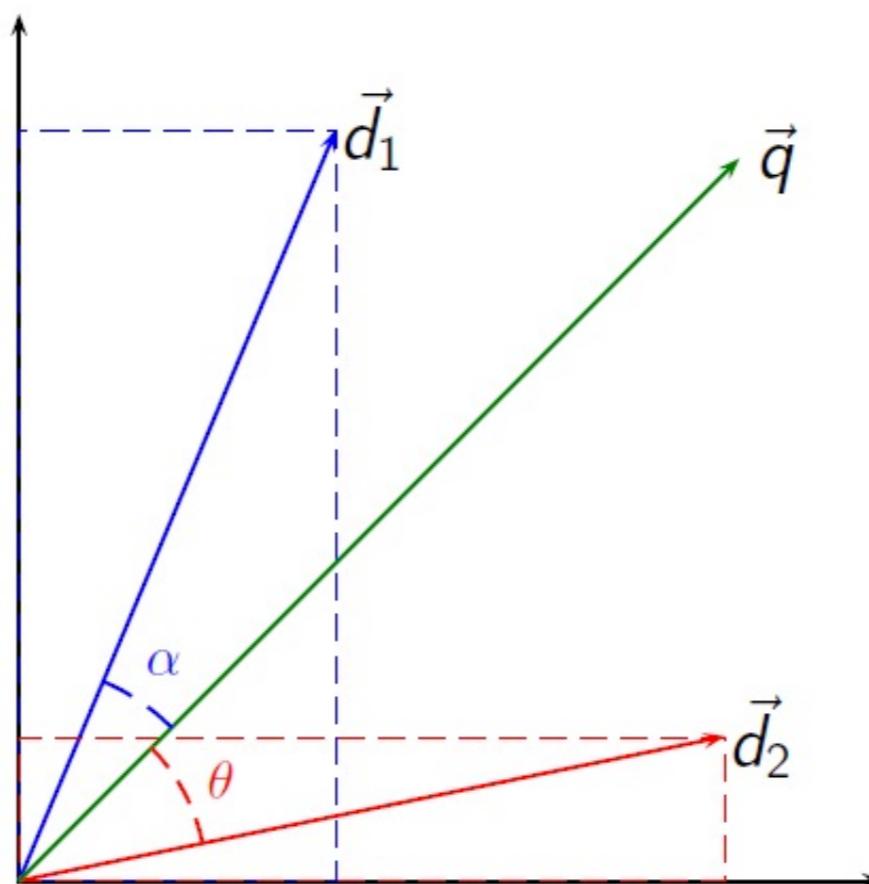
Species

Contains Peanuts

TYPES OF DATA

	Continuous	Categorical
Supervised	Regression	Classification
Unsupervised	Dimension Reduction	Clustering

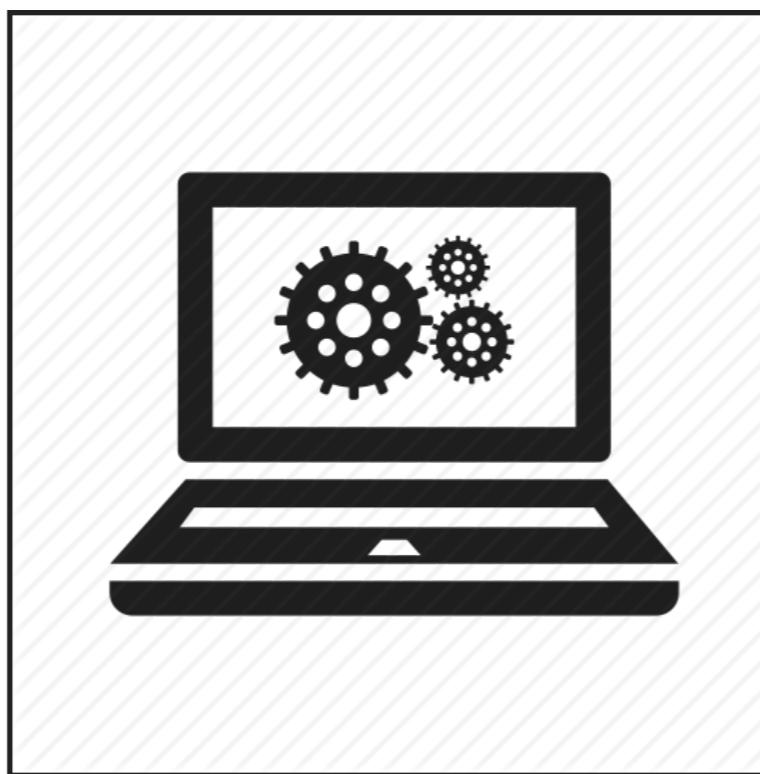
DATA LIVE IN A PLACE CALLED THE FEATURE SPACE. THIS IS A VECTOR SPACE.



EACH POINT IN THE SPACE IS CALLED A RECORD.

SUPERVISED LEARNING - WHAT'S IT DOING?

Input Data →



→

$f(x) \sim y$

SOME JARGON

Model Inputs: Features, Attributes, Predictors, Inputs, Independent Variables, Dimensions, probably more...

SOME JARGON

Model Inputs: Features, Attributes, Predictors, Inputs, Independent Variables, Dimensions, probably more...

Model Outputs (what we're trying to predict): Target, Response, Output, Dependent Variable, Labels

SOME JARGON

Model Inputs: Features, Attributes, Predictors, Inputs, Independent Variables, Dimensions, probably more...

Model Outputs (what we're trying to predict): Target, Response, Output, Dependent Variable, Labels

Row of Data (Inputs + Outputs): Observation, Datapoint, Record, Row

SOME JARGON

Model Inputs: Features, Attributes, Predictors, Inputs, Independent Variables, Dimensions, probably more...

Model Outputs (what we're trying to predict): Target, Response, Output, Dependent Variable, Labels

Row of Data (Inputs + Outputs): Observation, Datapoint, Record, Row

Labels: The values on the target variables in Supervised Learning

- ▶ **Feature Engineering** is the art of using your inputs in unique ways. So maybe you aren't getting good results with $y = m_1x_1 + m_2x_2 + b$ but can engineer $y = mx_1x_2 + b$.
- Function approximation + Feature engineering + Finding structure in the data = **Data Science.**

- ▶ **Supervised** is a 'game' of find the best way to map features to targets
 - Each data point is a vector in the feature space
 - There is (usually) only one target value (label)
- ▶ **Unsupervised** has no labels - we want to find structure
 - The task is to find the relationships between features